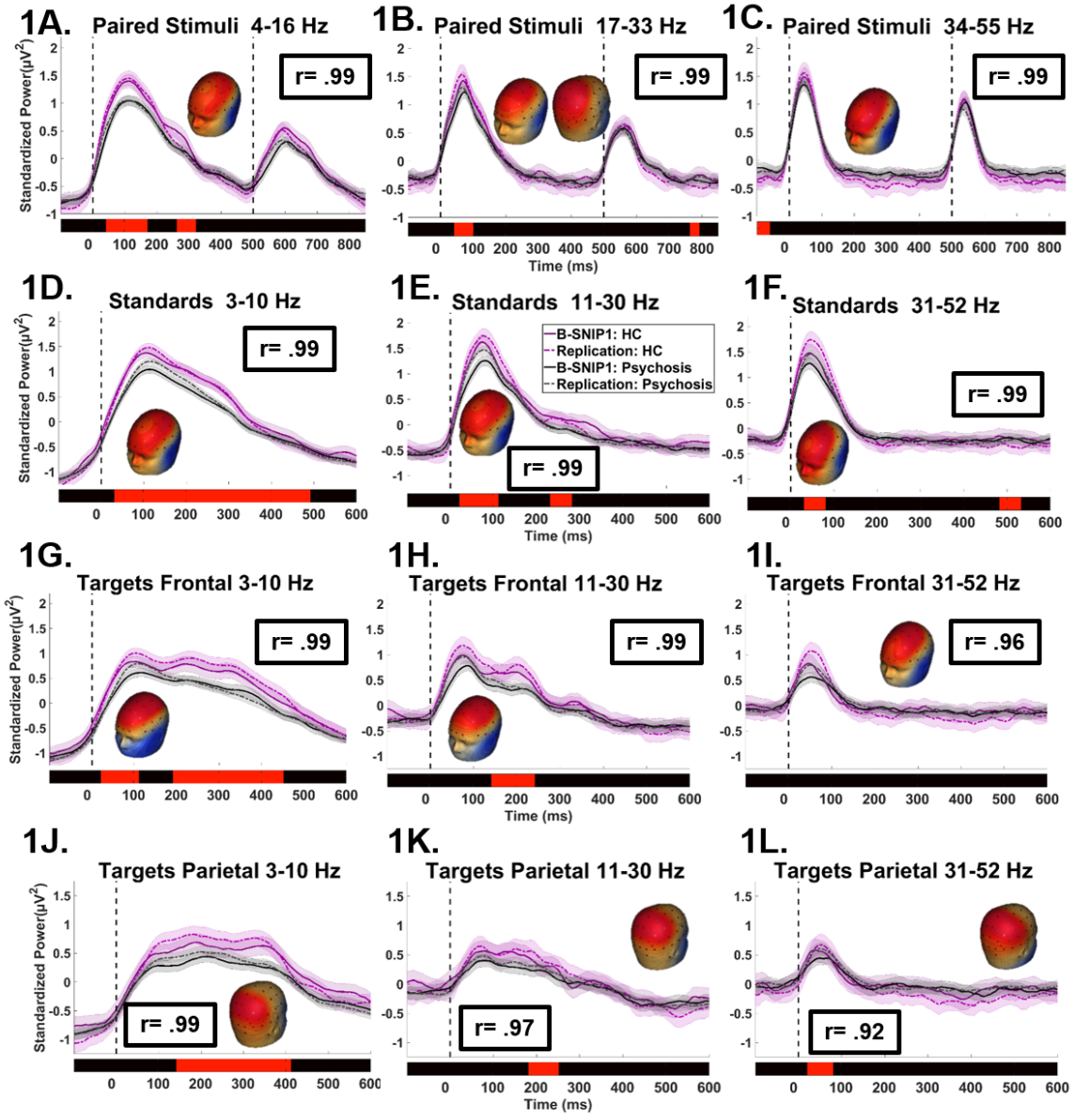
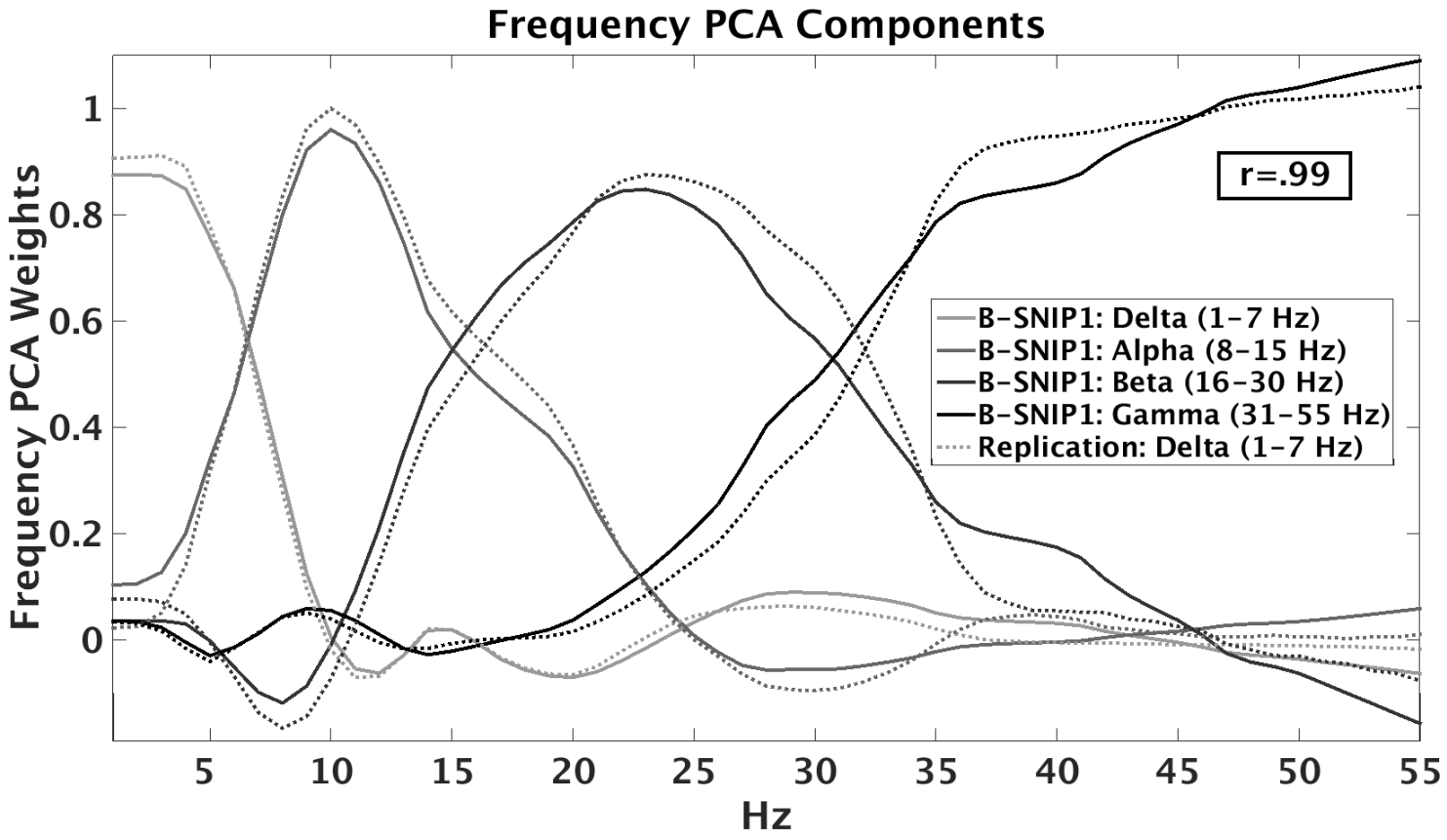
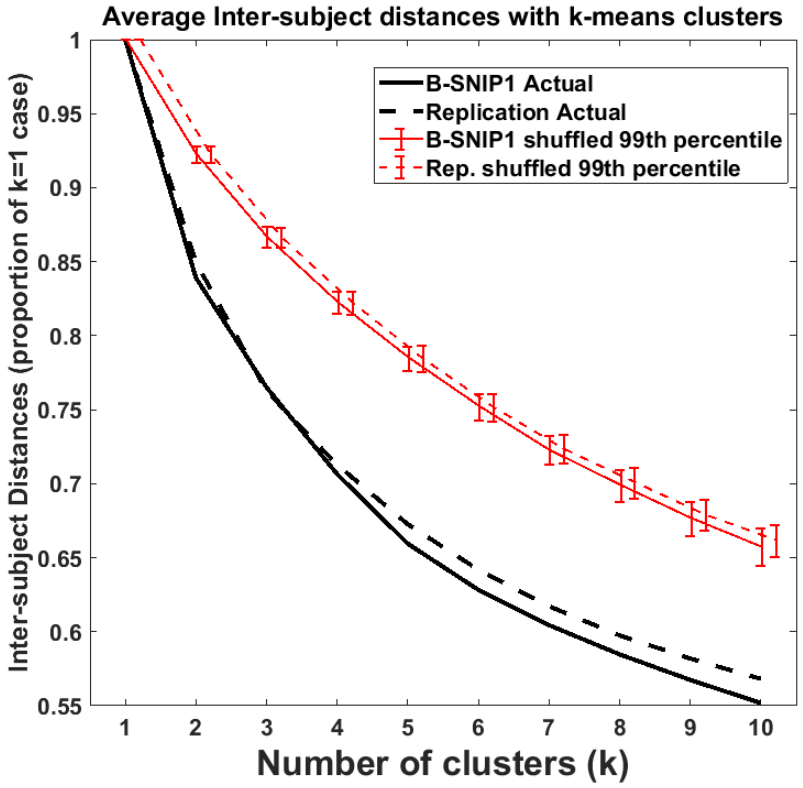
Supplementary Figure 1

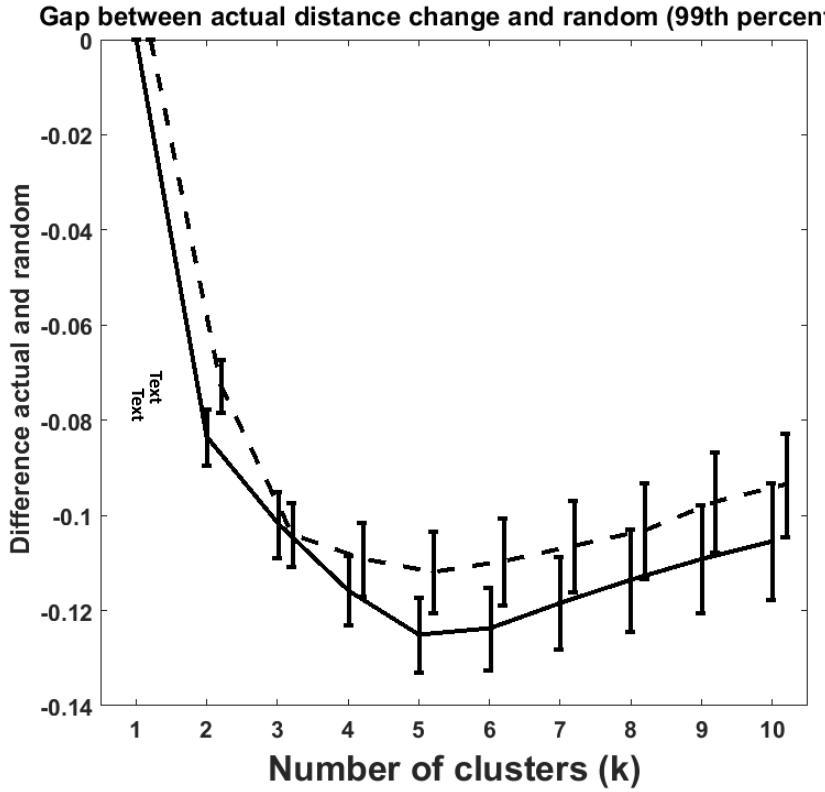


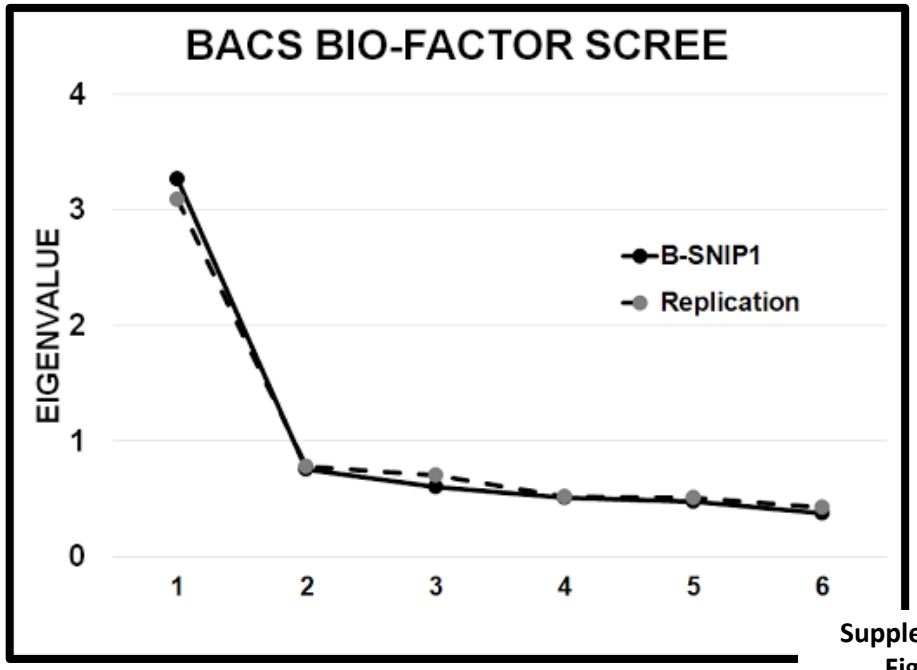
Supplementary Figure 2

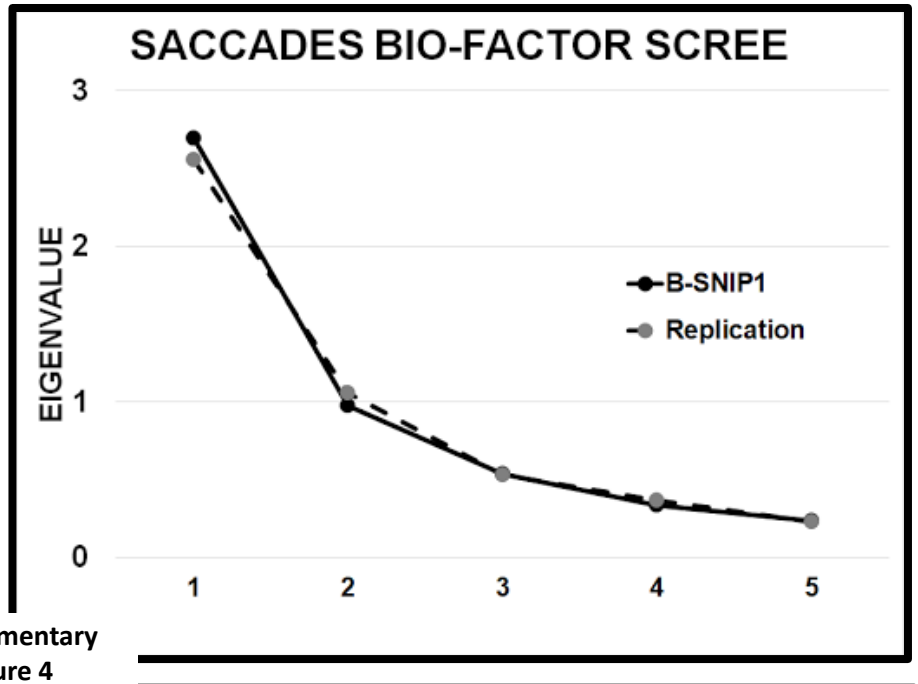


Supplementary Figure 3. Gap Statistic Results. The gap statistic provided one means for estimating the number of clusters in our data prior to implementing the k-means algorithm. The plot on the left shows differences between the pooled within-cluster sums of squares as a function of the number of requested clusters for the actual (black line) and null (red line) distributions. Functions are plotted for each distribution as a proportion of the 1 cluster (no subgroups) case. The solid line is for B-SNIP1 and the dashed line is for the replication sample. The plot on the right shows the gap function (the difference between the actual and null distribution functions), including the confidence intervals of the within cluster sums of squares at each requested cluster number. The gap outcome at the three cluster solution does not significantly differ from the 4-cluster case, so the three cluster solution most parsimoniously describes the number of subgroups given or data for both B-SNIP1 (solid line) and the replication samples (dashed line).

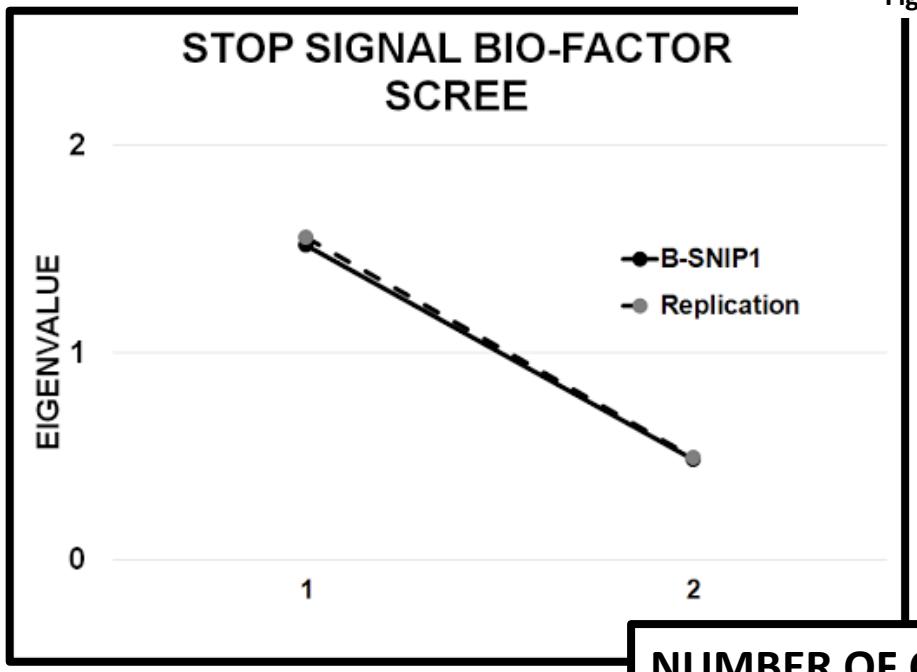


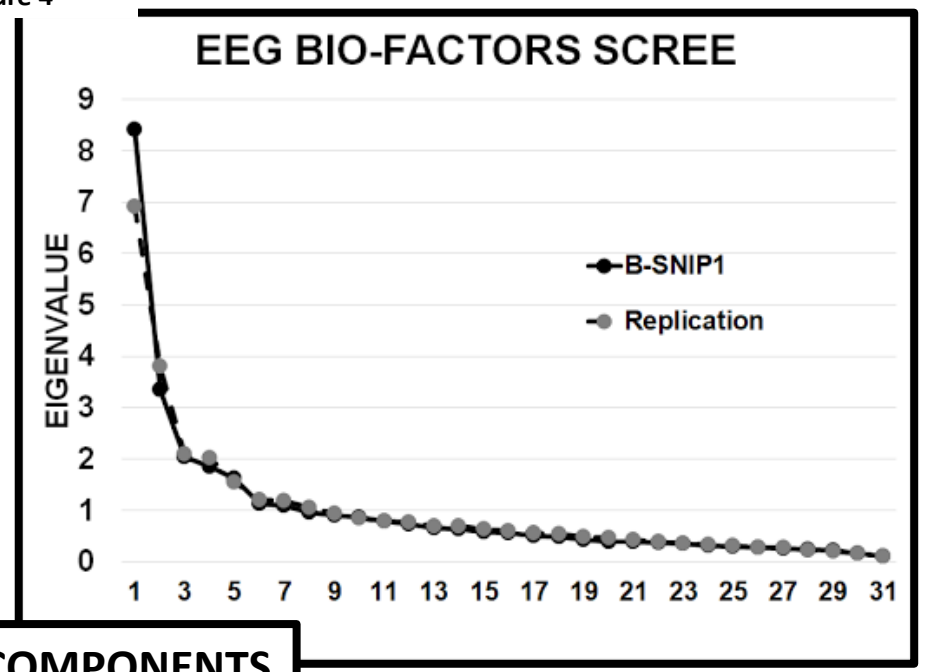






Supplementary Figure 4

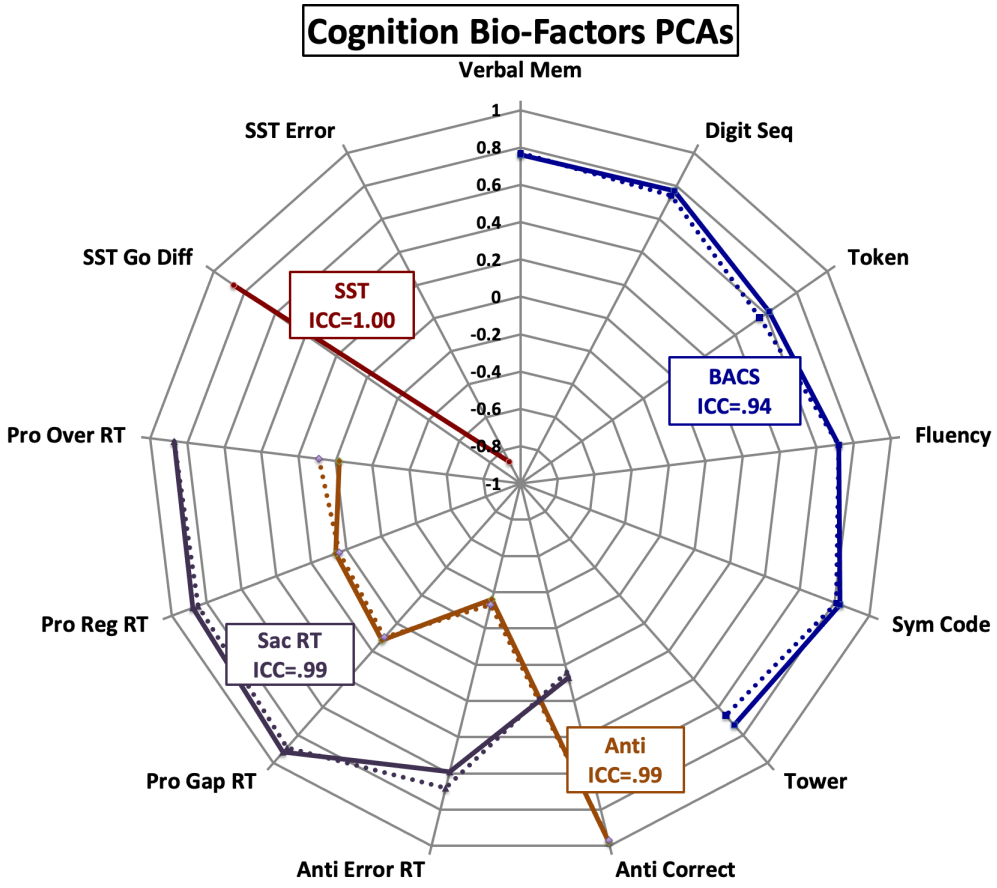




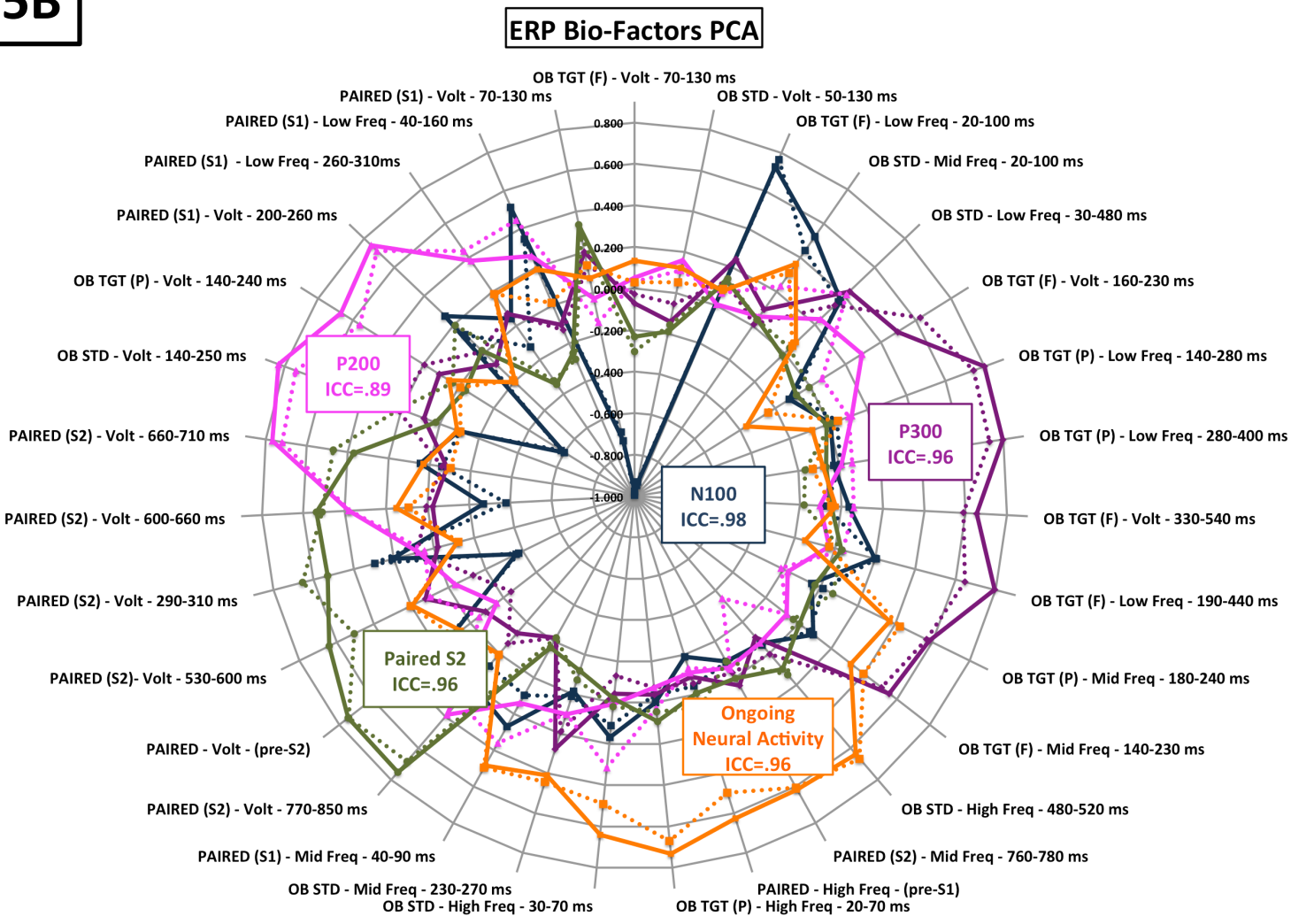
NUMBER OF COMPONENTS

Supplementary Figure 5. Bio-factor pattern similarity between B-SNIP1 and replication samples for Cognition and EEG/ERP variables. The patterns of similarity between B-SNIP1 and the replication samples on uniquely determined bio-factor patterns as derived from principal component analyses (PCA). The solid lines show the loadings for B-SNIP1 and the dotted lines show the loadings for the replication sample. The methods used here replicate those used previously. **(A).** Three different PCAs were run as a function of each variable set: one for the Brief Assessment of Cognition in Schizophrenia (BACS) variables (blue), one for the stop-signal task (SST) variables (maroon), and one for the combined anti- (Anti) and pro-saccade (Sac RT) variables (light brown and black, respectively). The plot shows the standardized loadings (from -1.0 to 1.0 on the principal axis) that are applied to each variable to obtain the bio-factor score from the multiple biomarkers. Variables with more extreme loadings contribute most to that particular bio-factor. Line and label colors highlight the different PCA outcomes, including the association (ICCs) between the projects on the PCA outcomes. There was one significant bio-factor for the BACS, one for the SST and two for the saccades. All four bio-factors showed high similarity between the projects. **(B).** One PCA was conducted on the 31 significant ERP variables, and this PCA revealed five significant bio-factors for each project (N100 ERP in dark blue, P300 ERP in purple, P200 ERP in magenta, ongoing neural activity in orange, and Paired S2 ERP in olive). The bio-factor loading values (from -1.0 to 0.9 on the principal axis) and labels are color-coded, and are displayed along with their individual associations between projects (ICCs). The five ERP bio-factors showed high similarity between projects. Inserted r-values are the correlations between the B-SNIP1 and replication samples. Variables with more extreme loadings contribute most to that particular bio-factor. Outside variable labels: OB indicates a variable came from oddball task, with STD and TGT indicating response to the OB standard and target stimuli, respectively, with F indicating the response with frontal cortex topography (P3a) and P indicating the response with the parietal cortex topography (P3b); PAIRED indicates a variable came from the paired-stimuli task, with S1, S2, pre-S1, and pre-S2 indicating variables were in response to the first, second, or preceded the first or second stimuli; Volt indicates the response was scored in the voltage-time domain, with Low, Mid, and High indicate the frequency range of the response scored in the frequency domain. Latency range of the variables (in msec) is also provided.

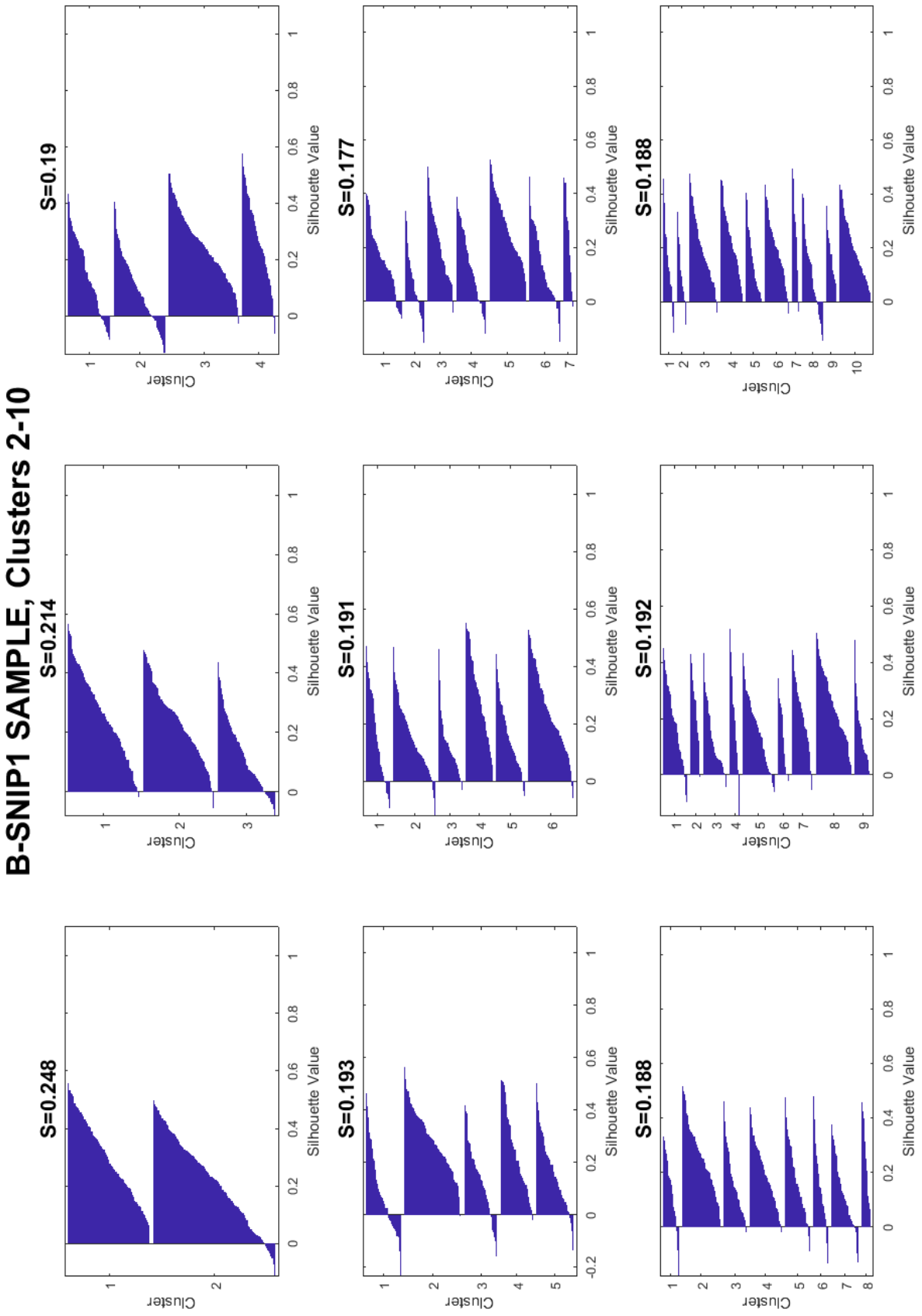
5A


Cognition Bio-Factors PCAs
Verbal Mem
Digit Seq
Token
Fluency
Sym Code
Tower
Anti Correct
Anti Error RT
Pro Gap RT
Pro Reg RT
Pro Over RT
SST Go Diff
SST Error
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
SST ICC=1.00
BACS ICC=.94
Sac RT ICC=.99
Anti ICC=.99

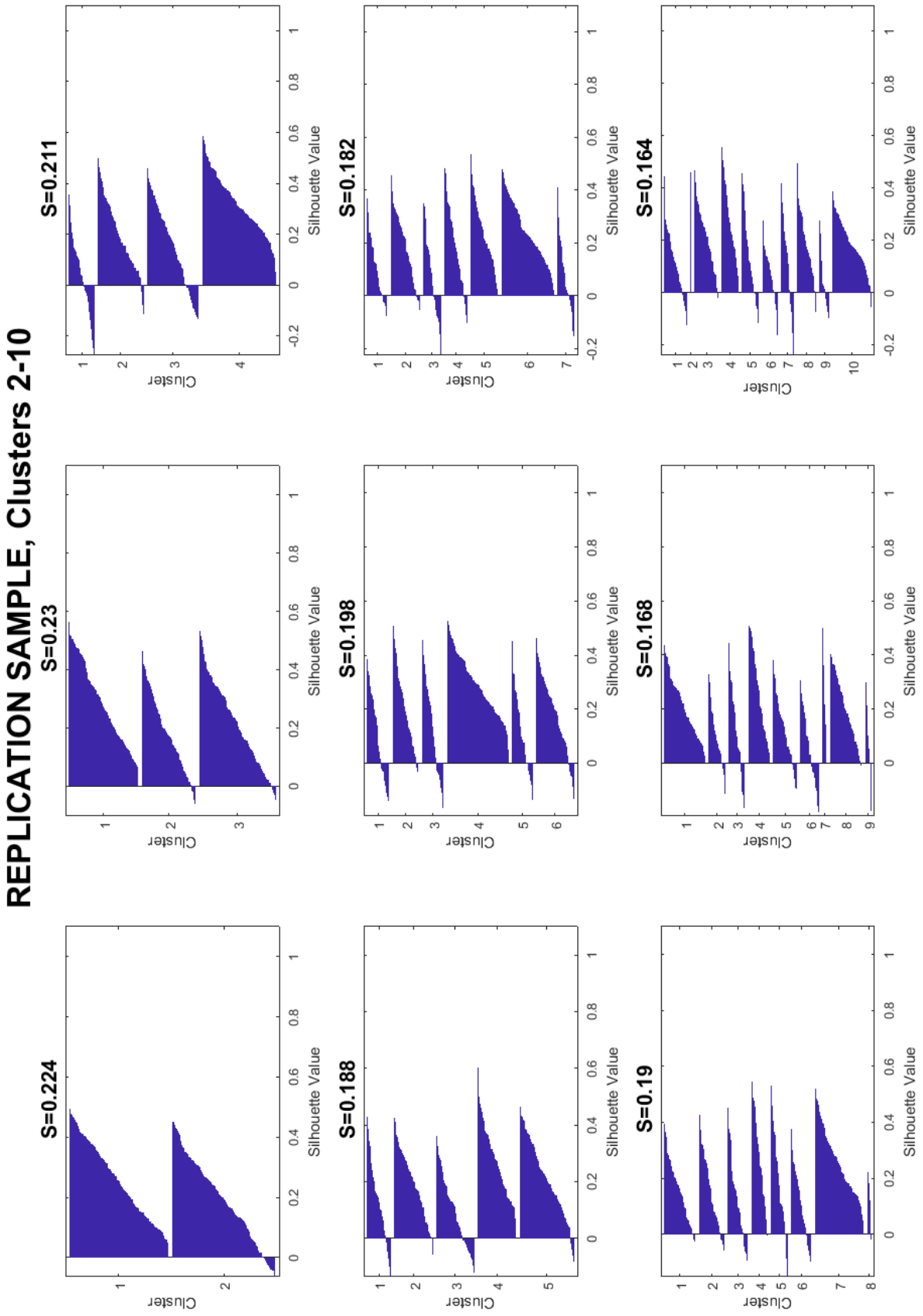
5B

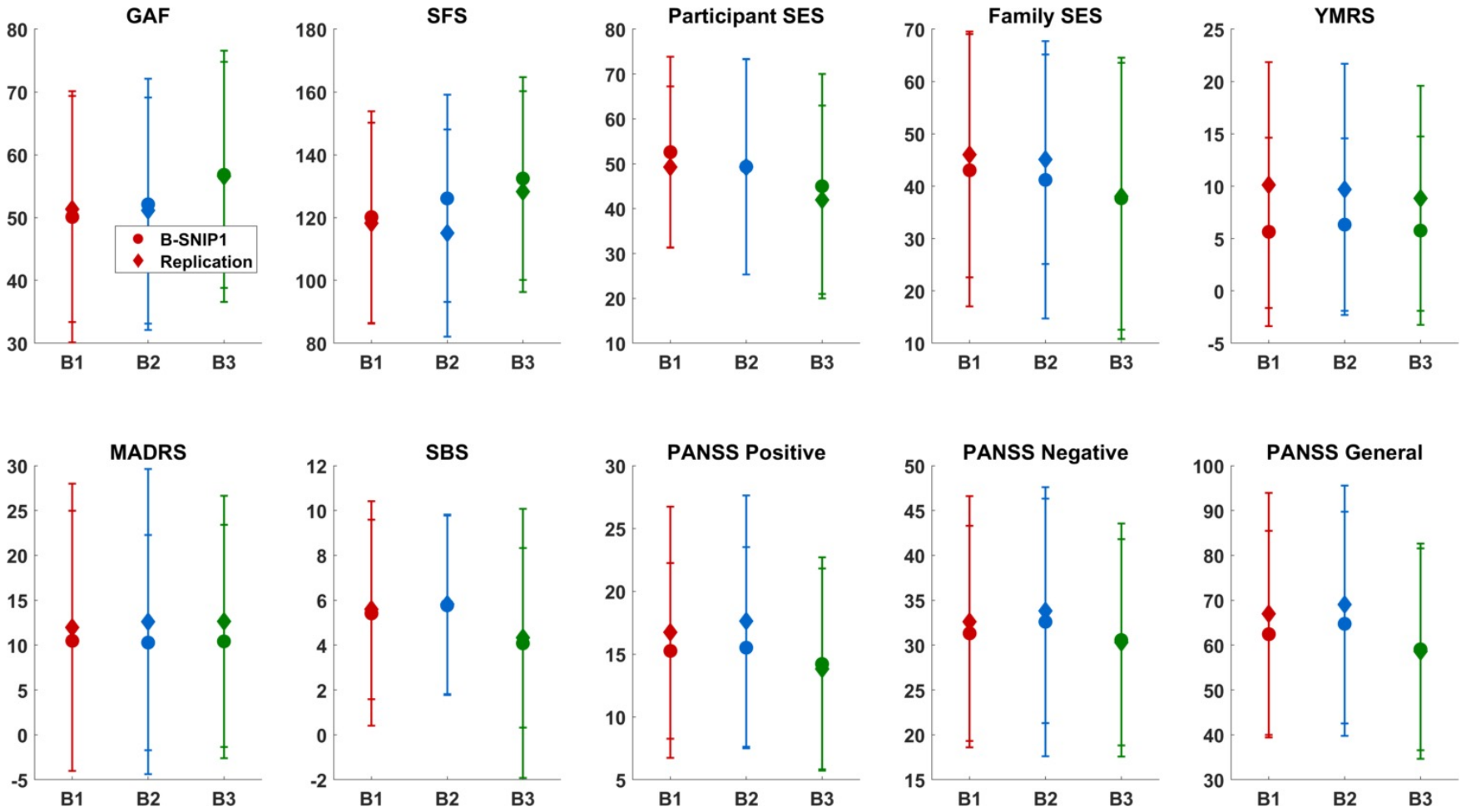

ERP Bio-Factors PCA
OB TGT (F) - Volt - 70-130 ms
OB STD - Volt - 50-130 ms
OB TGT (F) - Low Freq - 20-100 ms
OB STD - Mid Freq - 20-100 ms
OB STD - Low Freq - 30-480 ms
OB TGT (F) - Volt - 160-230 ms
OB TGT (P) - Low Freq - 140-280 ms
OB TGT (P) - Low Freq - 280-400 ms
OB TGT (F) - Volt - 330-540 ms
OB TGT (F) - Low Freq - 190-440 ms
OB TGT (P) - Mid Freq - 180-240 ms
OB TGT (F) - Mid Freq - 140-230 ms
OB STD - High Freq - 480-520 ms
PAIRED (S2) - Mid Freq - 760-780 ms
PAIRED - High Freq - (pre-S1)
OB TGT (P) - High Freq - 20-70 ms
OB STD - High Freq - 30-70 ms
OB STD - Mid Freq - 230-270 ms
PAIRED (S1) - Mid Freq - 40-90 ms
PAIRED (S2) - Volt - 770-850 ms
PAIRED - Volt - (pre-S2)
PAIRED (S2)- Volt - 530-600 ms
PAIRED (S2) - Volt - 290-310 ms
PAIRED (S2) - Volt - 600-660 ms
PAIRED (S2) - Volt - 660-710 ms
OB STD - Volt - 140-250 ms
OB TGT (P) - Volt - 140-240 ms
PAIRED (S1) - Volt - 200-260 ms
PAIRED (S1) - Low Freq - 260-310ms
PAIRED (S1) - Low Freq - 40-160 ms
PAIRED (S1) - Volt - 70-130 ms
0.800
0.600
0.400
0.200
0.000
-0.200
-0.400
-0.600
-0.800
-1.000
P200 ICC=.89
P300 ICC=.96
N100 ICC=.98
Paired S2 ICC=.96
Ongoing Neural Activity ICC=.96

Supplementary Figure 6

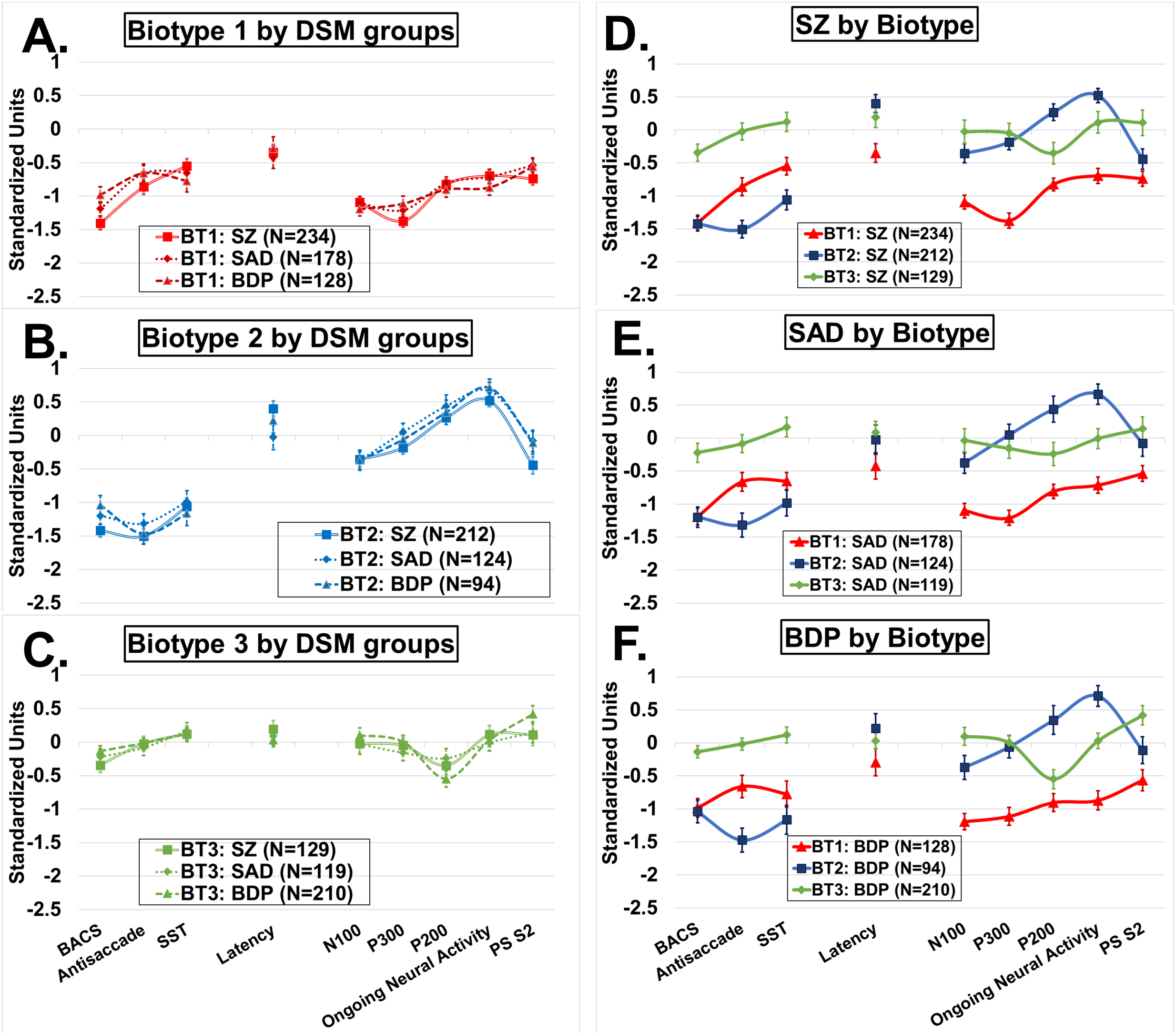


Supplementary Figure 7





Supplementary Figure 8. Clinical scales by Biotype and study. Average scores are plotted with interquartile range. Exact means, standard deviations, and statistical group comparisons are reported in Supplementary Tables 2 & 3.



Supplementary Figure 9: A-C: Average Biotype response for each Biofactor separated by DSM grouping. Error Bars represent a 90 C.I. to conservatively show overlap across DSM groups. **D-F**: Average DSM response for each Biofactor separated by Biotype grouping. Error bars represent a 95% C.I. to show the consistency of differences by Biotype.